**PROJECT PLAN**

**Dicoding Bootcamp - Capstone Project**

**INSTRUKSI**

1. **Hapus semua bagian deskripsi/instruksi** dalam setiap bab, termasuk instruksi ini.
2. Setiap grup hanya diperbolehkan untuk **submit project plan sebanyak 1x oleh salah satu anggota yang dijadikan perwakilan**.
3. ***Problem Statement*** adalah deskripsi permasalahan atau issue yang akan diselesaikan atau kondisi yang akan diperbaiki. Masukkan gap antara kondisi saat ini (problem) dengan tujuan akhir (terselesaikan) dalam bentuk proses atau produk. Umumnya hal ini diselesaikan atau berisi pertanyaan ***5W+1H (What, When, Why, Where, Who, How)***. Syarat pertama menyelesaikan permasalahan adalah mengerti dengan jelas kondisi dan latar belakang.
4. ***Research Question*** adalah pertanyaan atau hipotesis yang akan dibuktikan dalam percobaan/riset. Memilih research question yang tepat akan memudahkan Anda menyelesaikan riset. Sebuah penyelesaian masalah kemungkinan besar membutuhkan pengumpulan data, analisa, serta metode penyelesaian yang bervariasi. *Research Question* yang baik biasanya mencakup pertanyaan-pertanyaan yang sangat spesifik dan dalam. Contoh research question adalah sebagai berikut:
   1. Bagaimana orang lebih mudah memesan ojek? [Go-jek]
   2. Bagaimana orang lebih mudah belajar teknologi? [Dicoding]
   3. Bagaimana orang lebih mudah menemukan informasi dari seluruh dunia? [Google]
5. Perhatikan tata cara penulisan berikut:

**Font : Inter**

**Ukuran font : 12**

**Spasi : 1,5**

**Format file : .pdf**

1. Format penamaan file: **Project Plan - ID Grup**.pdf

**PROJECT PLAN**

**Dicoding Bootcamp - Capstone Project**

**ID Tim Capstone : DB4-PS009**

**Judul Proyek :** Mengidentifikasi Risiko Depresi pada Pelajar dengan Pembelajaran Mesin

**List Anggota :**

1. B244051F - Argo Wahyu Utomo - [**Aktif]**
2. B244021F - Bernadus Raka Sulistyo - [**Aktif]**
3. B244042F - Farah Dina - [**Aktif]**
4. B244044F - Wick Indiyarto - [**Aktif]**

**Ringkasan Eksekutif**

Depresi di kalangan pelajar adalah masalah yang semakin meningkat, yang dapat mempengaruhi kesehatan mental, performa akademik, dan kualitas hidup mereka secara keseluruhan. Saat ini, banyak mahasiswa yang mengalami tekanan akademik, pola tidur yang buruk, serta ketidakpuasan terhadap kehidupan mereka, yang berpotensi meningkatkan risiko depresi. Namun, banyak institusi pendidikan belum memanfaatkan data secara maksimal untuk mengidentifikasi mahasiswa yang berisiko tinggi mengalami depresi, padahal pendekatan berbasis data berpotensi meningkatkan upaya pencegahan lebih dini.

Melalui proyek ini, kami akan menggunakan **model klasifikasi** untuk memprediksi apakah seorang mahasiswa berisiko mengalami depresi berdasarkan faktor-faktor seperti tekanan akademik, pola tidur, kepuasan hidup, dan performa akademik. Dataset yang diberikan berisi label **1** (terdepresi) dan **0** (tidak depresi), dan model ini akan dilatih untuk mengidentifikasi mahasiswa yang mungkin membutuhkan perhatian lebih terkait kesejahteraan mental mereka. **Kondisi yang diharapkan** adalah terciptanya sistem yang dapat memprediksi secara akurat mahasiswa yang berisiko mengalami depresi, memungkinkan institusi pendidikan untuk merancang intervensi yang lebih tepat dan lebih awal. Dengan demikian, diharapkan dapat mengurangi angka depresi, meningkatkan kesejahteraan mahasiswa, serta memperbaiki performa akademik mereka.

**Cakupan Proyek dan Hasil Kerja**

Projek ini bertujuan untuk menjawab pertanyaan berikut:

1. Apa saja faktor utama yang berkontribusi terhadap risiko depresi pada mahasiswa?
2. Bagaimana hubungan antara tekanan kerja, tekanan akademik, dan pola hidup terhadap risiko depresi?
3. Bisakah model klasifikasi dibuat untuk mengidentifikasi mahasiswa yang berisiko tinggi mengalami depresi?

Hasil kerja berupa:

* Insight yang menjawab pertanyaan bisnis di atas berdasarkan analisis data.
* Model prediksi yang dapat mengklasifikasikan mahasiswa menjadi dua kategori: terdepresi (label 1) dan tidak terdepresi (label 0), berdasarkan faktor-faktor yang relevan.

**Jadwal Pengerjaan**

1. Minggu ke-1: Understanding the Problem
2. Minggu ke-2: Data Preparation
3. Minggu ke-3: Data Analysis & Feature Engineering
4. Minggu ke-4: Modeling
5. Minggu ke-5: Evaluation

**Sumber Daya Proyek**

Bahasa Pemrograman: Python

Library:

* pandas
* numpy
* matplotlib
* seaborn
* scikit-learn
* plotly (untuk visualisasi interaktif, opsional)

Dataset: <https://www.kaggle.com/datasets/hopesb/student-depression-dataset>

**Rencana Manajemen Risiko dan Isu**

1. Risiko dan Solusi

| **Risiko** | **Dampak** | **Strategi Mitigasi** |
| --- | --- | --- |
| Kualitas data buruk (missing values, outliers, dsb). | Analisis menjadi tidak akurat. | - Melakukan pembersihan data sejak awal.  - Menggunakan teknik imputasi data. |
| Kurangnya pemahaman awal terhadap masalah. | Solusi yang dihasilkan kurang relevan. | - Diskusi dengan tim untuk memastikan tujuan jelas.  - Mendefinisikan research question dengan baik. |
| Kesulitan dalam pemilihan model prediksi yang tepat. | Hasil prediksi tidak optimal. | - Mencoba beberapa algoritma dan membandingkan performanya.  - Tuning hyperparameter secara sistematis. |
| Kurangnya waktu untuk menyelesaikan proyek. | Proyek tertunda atau hasil tidak selesai. | - Membagi tugas secara efektif.  - Mengatur prioritas berdasarkan jadwal pengerjaan. |
| Masalah teknis seperti perangkat lunak atau alat analisis. | Menghambat pengerjaan proyek. | - Menyiapkan alat dan lingkungan kerja lebih awal.  - Backup data dan menggunakan versi stabil software. |
| Konflik dalam tim atau komunikasi yang buruk. | Kerja tim tidak efisien. | - Menjadwalkan pertemuan rutin.  - Menggunakan alat kolaborasi seperti Slack atau Trello. |

1. Analisis SWOT

| **Strengths (Kekuatan)** | **Weaknesses (Kelemahan)** |
| --- | --- |
| - Dataset mencakup banyak fitur yang relevan.  - Alur kerja yang terstruktur. | - Kualitas data awal mungkin memerlukan banyak pembersihan.  - Keterbatasan waktu pengerjaan. |
| **Opportunities (Peluang)** | **Threats (Ancaman)** |
| - Memberikan wawasan berharga untuk institusi pendidikan.  - Dapat digunakan untuk mengembangkan sistem deteksi dini. | - Risiko keterlambatan akibat kendala teknis.  - Ketidakpastian kualitas hasil analisis jika data kurang representatif. |